The Explicit Dependence of Quadrat Variance on the Ratio of Clump Size to Quadrat Size

Francis J. Ferrandino

Associate Scientist, Department of Plant Pathology and Ecology, The Connecticut Agricultural Experiment Station, P.O. Box 1106, New Haven 06504.

Accepted for publication 14 December 2004.

ABSTRACT

Ferrandino, F. J. 2005. The explicit dependence of quadrat variance on the ratio of clump size to quadrat size. *Phytopathology* 95:452-462.

In the past decade, it has become common practice to pool mapped binary epidemic data into quadrats. The resultant "quadrat counts" can then be analyzed by fitting them to a probability distribution (i.e., beta-binomial). Often a binary form of Taylor's power law is used to relate the quadrat variance to the quadrat mean. The fact that there is an intrinsic dependence of such analyses on quadrat size and shape is well known. However, a clear-cut exposition of the direct connection between the spatial properties of the two-dimensional pattern of infected plants in terms of the geometry of the quadrat and the results of quadrat-based analyses is lacking. This problem was examined both empirically and analytically. The empirical approach is based on a set of stochastically

generated "mock epidemics" using a Neyman-Scott cluster process. The resultant spatial point-patterns of infected plants have a fixed number of disease foci characterized by a known length scale (monodisperse) and saturated to a known disease level. When quadrat samples of these epidemics are fit to a beta-binomial distribution, the resulting measures of aggregation are totally independent of disease incidence and most strongly dependent on the ratio of the length scale of the quadrat to the length scale of spatial aggregation and to a lesser degree on disease saturation within individual foci. For the analytical approach, the mathematical form for the variation in the sum of random variates is coupled to the geometry of a quadrat through an assumed exponential autocorrelation function. The net result is an explicit equation expressing the intra-quadrat correlation, quadrat variance, and the index of dispersion in terms of the ratio of the quadrat length scale to the correlative length scale.

In disease epidemiology, binary epidemics are often characterized in terms of the number or fraction of infected plants within quadrats of a certain size. The average of these quadrat counts provides information on the incidence of disease, and the variation in disease incidence among the quadrats can be used to describe spatial aggregation (9,10,12). Such quadrat-based methods are most useful when the number of diseased plants in the field are too few to allow a full spatial correlation analysis (3,7,8). Recent orchard-based studies of the spatial analysis of diseases have used a series of different-sized quadrats (1,4) and present ample experimental evidence for the strong dependence of the quadrat variance on quadrat size. As Pielou observed (page 100 in literature citation 18), "...measures of aggregation based on quadrat data will not as a rule be unique; different values will be obtained with different quadrat sizes." Pielou goes on to state that there are two exceptions to the above general rule which occur when the quadrat size is either much larger or much smaller than what the author calls the intrinsic "granularity" of the spatial pattern, which translates to the length scale over which spatial correlation is important. Pielou's arguments were based on "count" data, and it is my purpose to extend these arguments to binary data. This reasoning suggests that it might be informative to express the intra-quadrat correlation or, equivalently, the quadrat variance in terms of the ratio of the length scale of a quadrat to the correlative length scale, hereafter denoted as a .

Previously, a two-dimensional stochastic spatial-contact simulation based on the half-Cauchy dispersal function has been used

to investigate the effect of initial infection, dispersal length scale, and quadrat size on quadrat-based statistics (26,27). These studies concluded "...differences among the values of these spatial statistics from different studies may reflect, at least in part, differences in quadrat sizes and initial conditions, and, to a lesser extent, differences in the spore-dispersal gradient." It is the last element of this conclusion, based on the relation between the dispersive length scale, quadrat size, and the resultant intra-quadrat correlation, that I would like to reexamine in more detail. In the above studies (26,27), the correlation was averaged over time and over all simulation variables which makes it impossible to tease out the effect of the ratio of the length scale of a quadrat to the correlative length scale.

Over the course of an epidemic, the effective correlative length scale will initially increase in time as disease foci spread in space. As these foci begin to overlap, the diseased field becomes an amalgam of patches of infected plants and relatively healthy areas forming an intricate patchwork quilt. The component patches will vary widely in physical size and in disease severity over the course of the epidemic. The size of these patches is initially governed by the dispersive length scale, but later in the epidemic, patch size is determined by the distance between foci. Thus, the correlative length scale will vary widely over time and will be strongly correlated with disease incidence and the distance between disease foci. This complication is exacerbated when the half-Cauchy dispersal distribution (26,27) is used to describe spore dispersal. Since this "fat-tailed" function has no defined first moment (6), spore dispersal in this case has no characteristic length scale. As discussed in Feller (6), an epidemic based on the half-Cauchy distribution results in the production of secondary satellite foci many times the median distance from the original infection. These secondary foci are created throughout the epidemic and so, at any one point in time, the spatial pattern of diseased plants is characterized by an ever-increasing spectrum of

Corresponding author: F. J. Ferrandino
E-mail address: Francis.Ferrandino@po.state.ct.us

disease foci having differing length scales and differing degrees of focal saturation.

I suggest an alternate approach using a series of spatial “test patterns” generated by a Neyman-Scott cluster process (3,8,16) to evaluate the effects of various length scales on quadrat-based analyses. Such patterns consist of randomly placed foci of diseased plants with well-defined length scales and levels of focal saturation. In this way, the length scales associated with quadrat pooling and the spatial correlation of disease, as well as overall disease incidence and the level of focal saturation can each be varied independently of one another. In analyzing these point-patterns, I intend to follow the path that Ridout and Xu (19) laid out as “the logical flow from spatial autocorrelation through intraclass correlation to quadrat variance.” In this paper, I first extend Pielou’s argument to binary quadrat data, then verify the resulting limits using a stochastic model, and finally derive an explicit equation expressing the intra-quadrat correlation in terms of a based on an assumed exponential form for the autocorrelation function.

THEORY AND APPROACHES

Neyman-Scott point-patterns. In order to test the impact of quadrat size, I first construct a series of spatial point-patterns or “mock epidemics” that vary in degree of aggregation and the length scale of spatial correlation. These two properties respec-

tively correspond to what Pielou calls the intensity and the granularity of the spatial pattern (18). The plan is to randomly generate a fixed number, F , of disease foci within a field containing N plants. Each of these foci are then used to generate infective offspring that are dispersed according to a given radial transport function, $T(r)$, of distance r from center of each focus. The details of the construction of these simulations for a 160×160 evenly spaced grid of plants are given in the Appendix. A simulation is totally determined given the following three parameters: (i) p , the probability of infection which is assumed equal to the number of infected plants within the field divided by the total number of plants; (ii) s , the average disease saturation within a focus which is equal to the number of infected plants within a focus divided by the total number of plants within the focus; and (iii) λ , the focal radius. The fixed number of foci, F , is then given by $F = Np/(s\pi\lambda^2)$.

For the following simulations, s is always chosen to be greater than p . With this in mind, the intensity of focal aggregation, $\aleph$, is defined by $\aleph = (s - p)/(1 - p)$, which varies from a value of zero to one as aggregation increases. In what follows, I denote a given simulation in terms of these parameters as the parenthetical triplet (p , s , and λ). Two such simulations ([0.3, 0.44, and 16] and [0.3, 1, and 4]) are illustrated in Figure 1A and B.

Pooled quadrats. We now consider the number of infected plants within contiguous 8×8 quadrats each containing 64 plants. I define L as the length of the side of a square quadrat in plant

Fig. 1. Results of spatial epidemic simulations within a 160×160 field of plants using two Neyman-Scott cluster processes with differing dispersive length scales (A, $p = 0.3$; $s = 0.44$; and $\lambda = 16$; B, $p = 0.3$; $s = 1$; and $\lambda = 4$, where p = fraction of plants infected, s = saturation of individual foci, and λ is the focal radius). Infected plants are shown as black squares. The results of pooling the 64 plants in each of the 400 8×8 quadrats are shown for each simulation (C and D) in 64 shades of gray corresponding to disease incidence within each quadrat.

spacing units. Obviously the number of infected plants within each quadrat, i , can vary discretely from 0 to a maximum value of $n = L^2$ (64 in this case). This is illustrated in Figure 1C and D where the 8×8 quadrats are shaded in 64 levels of gray. The average probability density distributions over 1,000 realizations for each of the pooled point-patterns illustrated in Figure 1C and D are plotted in Figure 2. If the diseased plants were distributed randomly, then the infected quadrat counts would be described by a binomial distribution with the expectation value for i equal to np and variance about this value equal to $np(1 - p)$ (Fig. 2). However, examining Figure 2 we see that the probability distributions obtained from the above examples have many more sparsely infected quadrats and highly infected quadrats than expected from random behavior. The observed counts are said to be “overdispersed.” The index of dispersion, ID, is defined as the ratio of the observed variance, σ_i^2 , to the binomial variance (i.e., $ID = \sigma_i^2/[np(1 - p)]$). The value of ID can range from 0 for homogeneous distributions to a maximum value of n for totally aggregated distributions. Such overdispersion can, alternatively, be modeled using the beta-binomial distribution (10; Appendix).

The beta-binomial distribution. The beta-binomial probability distribution has two parameters which vary from zero to unity: mean incidence, p , and the inter-quadrat correlation coefficient, κ . When the value of κ approaches zero, the beta-binomial distribution approaches the binomial distribution indicative of a random spatial arrangement. As the value of κ increases so does overdispersion, indicating increased spatial aggregation. Alternatively, the beta-binomial aggregation parameter, $\theta = \kappa/(1 - \kappa)$, which varies from zero to infinity, can be used as an aggregation index. The above parameters can be fit by the method of moments (9–11,15) so that

$$\kappa = \frac{ID-1}{n-1} = \frac{\left[\frac{\sigma_i^2}{np(1-p)} - 1 \right]}{n-1} \quad (1)$$

where σ_i^2 is the variance of the infected quadrat counts, p is the mean incidence (equation A2), and n is the number of plants (branches, leaves, or leaflets) contained within each quadrat. For the two data sets shown in Figure 1, equation 1 yields $\kappa = 0.197$

and $\kappa = 0.182$ for $\lambda = 16$ (Fig. 1A and C) and $\lambda = 4$ (Fig. 1B and D), respectively. When these simulations are repeated 1,000 times, the average value of κ (equation 1) approaches the same value (0.186) for both sets of parameters (i.e., p , s , and λ). Also shown are the beta-binomial fits (method of moments) to the data.

The above analysis was repeated using $L = 2, 4, 16$, and 32 in addition to the original $L = 8$ plant spacings on a side. The average calculated values for κ using 1,000 simulations are plotted versus quadrat size in Figure 3. Obviously, a quadrat-based evaluation of the relative aggregation of these two stochastic processes strongly depends on quadrat size. This point is further illustrated in Figure 4 where fitted values of κ are plotted versus L for all of the simulations. The 400 values of κ are the averages over 1,000 simulations for 80 different combinations of epidemic pa-

Fig. 3. The relation between the intra-quadrat correlation coefficient, κ , and quadrat length scale for the two spatial epidemic simulations shown in Figure 1A, $[(p, s, \text{ and } \lambda) = (0.3, 0.44, \text{ and } 16)]$ and B, $[(p, s, \text{ and } \lambda) = (0.3, 1, \text{ and } 4)]$, where p = fraction of plants infected, s = saturation of individual foci, and λ is the focal radius. Note that an increased saturation of disease within a focus, s , shifts the plot upward, and the slope of the plot decreases with an increase in the dispersive length scale, λ .

Fig. 2. Probability distributions for 1,000 realizations of the spatial epidemic simulations illustrated in Figure 1A and B. Also shown are the almost identical beta-binomial fits to the data sets (BBD4 and BBD16) and the binomial distribution (Binom) for comparison.

rameters ($p = 0.075, 0.15, 0.3$, and 0.6 ; $\aleph = (s - p)/(1 - p) = 0.2, 0.4, 0.6, 0.8$, and 1 ; and $\lambda = 2, 4, 8$, and 16) sampled using five different quadrat sizes ($L = 2, 4, 8, 16$, and 32).

A simple self-similarity transformation. It is reasonable to assume that the intensity of aggregation, characterized by κ , will approximately scale with the average disease saturation within a fo-

cus, s , and that any length scale-dependence of this variable will depend on the ratio L/λ . Using this reasoning, I plot κ/s versus L/λ in Figure 5 using the same data plotted in Figure 4. This simple transformation, which is totally independent of disease incidence, accounts for a great deal of the variation in the observed value of κ . If we look at log-log plot of this data (Fig. 5, inset), it

Fig. 4. Variation in the value of the intra-quadrat correlation, κ , plotted as a function of quadrat length scale, L , for simulated spatial epidemics using a Neyman-Scott cluster process. The 400 values of κ are the averages over 1,000 simulations for 80 different combinations of epidemic parameters ($p = 0.075, 0.15, 0.3$, and 0.6 ; $\aleph = (s - p)/(1 - p) = 0.2, 0.4, 0.6, 0.8$, and 1 ; and $\lambda = 2, 4, 8$, and 16) sampled using five different quadrat sizes ($L = 2, 4, 8, 16$, and 32). Legend labels refer to the triplet (p, s , and λ), where p = fraction of plants infected, s = saturation of individual foci, and λ is the focal radius.

Fig. 5. The ratio κ/s plotted versus relative length scale L/λ for the simulated spatial epidemics using a Neyman-Scott cluster process illustrated in Figure 4. Inset shows a log-log plot of the same data. The 400 values of κ are the averages over 1,000 simulations for 80 different combinations of epidemic parameters [$p = 0.075, 0.15, 0.3$, and 0.6 ; $\aleph = (s - p)/(1 - p) = 0.2, 0.4, 0.6, 0.8$, and 1 ; and $\lambda = 2, 4, 8$, and 16] sampled using five different quadrat sizes ($L = 2, 4, 8, 16$, and 32). Straight lines illustrate Pielou's (18) predictions. Dashed line was calculated using equation 12 with $\rho(0) = 0.5 s$. In the above simulations, κ = the intra-quadrat correlation, p = fraction of plants infected, s = saturation of individual foci, λ is the focal radius, and $\aleph$ = the intensity of focal aggregation.

is obvious that the κ remains relatively independent of quadrat size when L/λ is <1 . However, when the quadrat is larger than a disease focus ($L/\lambda > 1$), the value of κ approaches an inverse power law of the ratio, L/λ . Both of these results have already been presaged by Pielou over 35 years ago (18).

Pielou's two exceptions. After making the broad statement that measures of aggregation will depend on quadrat size which I have already quoted, Pielou goes on to outline two special types of spatial pattern for which suitably chosen measures of aggregation will be independent of quadrat size (page 100 in literature citation 18). Pielou was concerned with count data and uses a Poisson distribution to make the arguments. In what follows and in a detailed discussion in the Appendix, I paraphrase Pielou's exposition to reflect the binary nature of the problem at hand for which it is appropriate to use a binomial distribution to describe random behavior.

Pattern 1. Consider a population of well-separated small foci of disease so widely spaced that they are rarely cut by the edges of a quadrat and do not interfere with one another. In this case the number of infected plants in a quadrat is simply the product of the number of included foci and the number of infected plants within each focus, i_F . As is shown in the Appendix, the result of this clumping of infected plants is that the variance in the observed number of infected plants per quadrat is equal to the binomial variance, $np(1-p)$, multiplied by the number of infected plants in a focus, i_F . The quantity i_F is approximately equal to the product of the area of a focus and the saturation of disease within the focus ($i_F = s \times \pi\lambda^2$; equation A12). Substituting equation A12 into equation 1, we get

$$\kappa = \frac{i_F - 1}{n - 1} = \frac{s\pi\lambda^2 - 1}{L^2 - 1} \quad (2')$$

This equation describes the asymptotic behavior of κ as $L \gg \lambda$, which implies $i_F < L^2$. In this case, an upper bound for the value of κ/s can be obtained by ignoring the subtraction of 1 from the numerator and denominator in equation 2' and dividing both sides by s , yielding

$$\frac{\kappa}{s} < \pi \left(\frac{L}{\lambda} \right)^{-2} \quad (2)$$

which yields a straight line with a slope of -2 on a log-log graph of κ/s versus L/λ (Fig. 5, straight line in inset). A direct comparison of the first equalities in equations 1 and 2' reveals that the index of dispersion, ID, is equal to the number of diseased plants in each focus, i_F , when $L \gg \lambda$.

Pattern 2. Next, consider a mosaic pattern comprised of M relatively large patches (much larger than a quadrat) over which disease incidence is relatively constant. In this case, we assume that the patches are so much larger than the quadrats that we can neglect quadrats that straddle the boundaries. In this case the variance is given by (equation A16)

$$\sigma_i^2 = np(1-p) \cdot \left[1 + (n-1) \frac{\sigma_p^2}{p(1-p)} \right] \quad (3)$$

where

$$\sigma_p^2 = \sum_{j=1}^M \frac{m_j}{N} (p_j - p)^2$$

where p_j is the disease incidence in the j th patch which includes m_j plants. A comparison of equation 3 and equation 1 leads us to the following identification:

$$\kappa = \frac{\sigma_p^2}{p(1-p)} \quad (4)$$

Thus, κ is independent of quadrat size. This behavior is reflected in the flat behavior of the κ/s curve in Figure 5 for $L \ll \lambda$. Note that the denominator in equation 4, $p(1-p)$, is the variance one would obtain if a fraction, p , of the area had a probability of being infected of unity and a fraction, $(1-p)$, had a probability of being

infected of zero. This, of course, is the maximum possible variance so that $\kappa \leq 1$.

In addition to providing a somewhat heuristic explanation of the geometry of the curve in Figure 5, equation 4 signals the implicit connection between intra-quadrat correlation and inter-quadrat variation. However, the spatial point-patterns used in the above analysis are constructs. As such, they are monodisperse and homogenous. Thus, the disease foci are characterized by a constant and well-defined dispersal length scale and each focus is associated with the same number of infected plants. In the real world, this is rarely the case and the simple arguments made above do not necessarily apply. In order to examine the effect of quadrat size in the general case, we need to formalize the statistics involved in taking quadrat counts.

The statistics of quadrat counts. Assume a field containing N plants planted in square array is being sampled using square quadrats L plant spacings on a side. The probability of infection is p , and the observed number of infected plants in a quadrat is i . Define the binary variable X_j which takes on a value of either 0 or 1 depending on whether or not the j th plant is healthy or infected. The total number of plants in each quadrat is $n = L^2$, and i is defined as

$$i = \sum_{j=1}^n X_j \quad (5)$$

The expectation value for i is simply the sum over the expectation values, $E(X_j) = p$, for each plant, so that $E(i) = np$. If the X_j 's are mutually independent, then the variance in i , written as σ_i^2 , would simply be the sum of the individual variances [$\sigma_i^2 = np(1-p)$, the binomial variance]. For the general case where there is interdependence, σ_i^2 is given by

$$\sigma_i^2 = \sum_{j=1}^n \sigma_{X_j}^2 + 2 \cdot \sum_{j>k} \text{cov}(X_j, X_k) \quad (6)$$

$$\sigma_i^2 = np(1-p) \cdot \left[1 + \frac{2}{n} \sum_{j>k} \rho(X_j, X_k) \right] \quad (7)$$

where $\text{cov}(\cdot)$ is the covariance and $\rho(\cdot)$ is the correlation coefficient between the two points in question. Of course, if the X_j 's are mutually independent, then the ρ 's in equation 7 are zero and the result is the binomial variance. On the other hand, if every pair of plants within a quadrat is perfectly correlated, then all the ρ 's are 1. This means that if one plant in a quadrat is infected, then all are infected and vice versa. In this case, since the number of unique pairs of plants for a quadrat containing n plants is $n(n-1)/2$, then equation 7 attains the maximum possible value which is indicative of total aggregation: $\sigma_i^2 = n^2 p(1-p)$. A simple comparison of equation 7 with equation 1 leads to the following identification of κ (2,19):

$$\kappa = \frac{2 \sum_{j>k} \rho(X_j, X_k)}{n(n-1)} \quad (8)$$

which is nothing more than the average value of the correlations over every possible pair of plants within a quadrat. Assume that the correlation function in equation 8 depends only on the radial distance between plant centers. Then the calculation of the above summation can be considerably simplified if we know the fraction of the total plant-pairs which are a certain distance apart. Defining the intra-quadrat point-pair radial probability distribution, $\phi_L(r_{jk})$, to be the fraction of the total number of plant-pairs from within a square quadrat with L plants on a side, which are a certain distance r_{jk} apart, equation 8 becomes (equation A19):

$$\kappa = \sum_{j=1}^{L-1} \sum_{k=0}^j \phi_L(r_{jk}) \rho(r_{jk}) \quad (9)$$

where $r_{jk} = (j^2 + k^2)^{1/2}$. All the possible combinations of j and k for a 3×3 quadrat ($L = 3$) are shown in Table 1. The algorithm for calculating $\phi_L(r_{jk})$ is outlined in the Appendix.

The continuous form of $\phi_L(r)$. The first and second moments (mean, variance, and standard deviation) for $\phi_L(r)$ for $L = 2$ to 8, 16, and 100 are shown in Table 2. As L grows larger, the mean of the distribution rapidly approaches a value close to $L/2$ and the standard deviation approaches a value near $L/4$. Due to the discrete nature of the ϕ_L function and the fact that the number of allowed distances increases rapidly with L , plots of this distribution are not very informative. For this reason, I define the cumulative pair distribution, $\Phi_L(r)$ (Appendix), as the fraction of the intra-quadrat plant-pairs that are less than or equal to a distance r apart, which is plotted versus r/L in Figure 6. As L increases, Φ_L rapidly approaches a smooth sigmoidal distribution centered about $L/2$ with a variance approaching $L^2/16$.

Of course, plants are not geometrical points; their foliage is distributed over an area. In fact, intra-plant distances are often comparable to inter-plant distances. Assuming a homogenous distribution of foliage throughout the field, the discrete function, $\phi_L(r_{jk})$, becomes the continuous radial density function, $\phi(r/L)$ (Appendix):

$$\phi(r/L)d(r/L) = \begin{cases} \frac{r}{L} \left[2\pi - 8\left(\frac{r}{L}\right) + 2\left(\frac{r}{L}\right)^2 \right] & ; r \leq L \\ \frac{r}{L} \left[2\pi - 4 - 8\cos^{-1}\left(\frac{L}{r}\right) + 8 \cdot \sqrt{\left(\frac{r}{L}\right)^2 - 1} - 2\left(\frac{r}{L}\right)^2 \right] & ; r > L \end{cases} \quad (10)$$

Equation 10 is identical to the result obtained in the limit as $L \rightarrow \infty$ (25). The moments of equation 10 can be calculated by direct integration (Table 2). If we now assume that the correlation function can be expressed as a negative exponential function of length scale, λ , equation 9 can be rewritten as

$$\kappa = \frac{1}{L} \int_0^{\sqrt{2}L} \phi(r/L) \rho(r) dr \quad \text{where} \quad \rho(r) = \rho(0)e^{-r/\lambda} \quad (11)$$

and $\phi(r/L)$ is given by equation 10. The constant, $\rho(0)$, is the maximum value of the autocorrelation function. Setting $a = L/\lambda$, the integral in equation 11 can be expressed as

$$\kappa = \rho(0) \left[\frac{2\pi}{a^2} - \frac{16}{a^3} + \frac{12}{a^4} - \frac{2e^{-a}}{a^4} \left((\pi-3)a^3 - (5-\pi)a^2 - 2a + 6 \right) + \left(\frac{19}{6} - \pi \right) e^{-a} \right] \quad (12)$$

The first four terms in equation 12 are the result of a direct integration of equation 11 ($r \leq L$) and account for more than

Fig. 6. Fraction of infected plant-pairs less than or equal to a certain distance, r_{jk} , apart $[\Phi_L(r_{jk})]$ plotted versus normalized distance (r_{jk}/L) for square quadrats L plants on a side. The median value $[\Phi_L(r_{jk}) = 0.5]$ occurs for pair separations about half the quadrat size apart ($r_{jk} = L/2$; Table 2). The solid line is the asymptotic limit ($L \rightarrow \infty$) and was obtained from a direct integration of equation 10.

TABLE 1. Possible point-pair distances and the corresponding Δx and Δy displacements for spatial epidemic simulations in a 3×3 quadrat of plants^a

$r_{jk} = (\Delta x^2 + \Delta y^2)^{0.5}$	j	k	$(\Delta x, \Delta y)$ $(-\Delta x, -\Delta y)$	$(L-j)(L-k)^b$ $L = 3$	α^c	Number of unique pairs
1	1	0	(+1,0), (0,+1) (-1,0), (0,-1)	6	2	12
$2^{1/2}$	1	1	(+1,+1), (+1,-1) (-1,-1), (-1,+1)	4	2	8
2	2	0	(+2,0), (0,+2) (-2,0), (0,-2)	3	2	6
$5^{1/2}$	2	1	(+2,+1), (+2,-1), (+1,+2), (+1,-2) (-2,-1), (-2,+1), (-1,-2), (-1,+2)	2	4	8
$8^{1/2}$	2	2	(+2,+2), (+2,-2) (-2,-2), (-2,+2)	1	2	2

^a In general, for an $L \times L$ quadrat containing $n = L^2$ plants, there are $L(L+1)/2 - 1$ unique radial distances between plants. For the case where $L = 3$ there are five possible point-pair distances corresponding to all the possible Δx and Δy displacements within a 3×3 quadrat. In this notation, plant 1 is at location $r_1 = (x_1, y_1)$, plant 2 is at location $r_2 = (x_2, y_2)$, $\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$, $j = \max(|\Delta x|, |\Delta y|)$, and $k = \min(|\Delta x|, |\Delta y|)$.

^b The product of $(L-j)$ and $(L-k)$ is the number of plant-pairs within an $L \times L$ quadrat that are separated by distance vector (j,k) (7).

^c α represents the degree of degeneracy due to symmetry.

TABLE 2. Calculated spatial moments for the intra-quadrat point-pair radial probability distribution, $\Phi_L(r)$ (equation A18), as a function of quadrat size (L) for spatial epidemic simulations

L	Mean	Variance	Mean/ L	Variance/ L^2	Standard deviation/ L
2	1.14	0.038	0.569	0.0095	0.098
3	1.63	0.327	0.545	0.0363	0.191
4	2.14	0.746	0.535	0.0466	0.216
5	2.65	1.291	0.531	0.0516	0.227
6	3.17	1.962	0.528	0.0545	0.233
7	3.68	2.757	0.526	0.0563	0.237
8	4.20	3.675	0.525	0.0574	0.240
16	8.36	15.47	0.522	0.0604	0.246
100	52.14	614.4	0.521	0.0614	0.248
Equation 10 ^a			$0.519 = 47/30 - \pi/3$	$0.0635 = 1/3 - (47/30 - \pi/3)^2$	$0.252 = [1/3 - (47/30 - \pi/3)^2]^{1/2}$

^a This row was obtained by direct integration of equation 10.

97.5% of the value of κ . The fifth term is an approximation of the effect of the corners of the quadrat ($r > L$), and its inclusion lowers the error of the prediction to less than 0.2%. As written, equation 12 provides an excellent estimate of the quadrat size-dependence of κ with the assumption that $\rho(0) = s/2$ (plotted in the inset of Figure 5 as a dashed line). Note that in the asymptotic limit as $a \rightarrow \infty$ (large quadrat size), equation 12 is consistent with equation 2 and as $a \rightarrow 0$ (small quadrat size), $\kappa \rightarrow \rho(0)$ and is relatively independent of L .

RESULTS AND DISCUSSION

For the 80 combinations of the parameters (p , s , and λ) used to generate Figure 5, the values of the correlation coefficient $\rho(r)$ have been calculated elsewhere (8). Since $\rho(r)$ scales with s and r is logically scaled by λ , the calculated values of $\rho(r)/s$ are plotted versus r/λ in Figure 7 for 8 of the 80 sets of epidemic parameters used. As s increases there is a reduced correlation at small length scales ($r < \lambda$) and a slight increase in correlation for $r > \lambda$ due to the effects of saturation. However, in general, $\rho(r)$ is determined by the values of the two parameters s and λ and is relatively insensitive to changes in the overall disease incidence, p . When these empirical $\rho(r)$ functions are inserted into equation 8 the result is very similar to Figure 5. The moment-based κ calculated from equation 1 versus the correlation-based κ calculated using equation 9 is plotted in Figure 8. A regression of the data from equation 1 on the predictions given by equation 12 results in regression parameters that are not significantly different from those obtained in Figure 8. For this reason the second, almost identical graph (variation $< 1\%$), is not presented.

The derivation of equation 12 (equations 5 to 12, equations A21 to A23) was obtained following Ridout and Xu's (19) "logical flow from spatial autocorrelation through intraclass correlation to quadrat variance." However, by passing to the continuous limit and putting off numerical evaluation until the end of the analysis, we have gained an insight into the role of geometry in determining the value of κ . Equation 12 is explicitly dependent on only two factors: $\rho(0)$ and $a = L/\lambda$. Thus, one single plot of $\kappa/\rho(0)$ versus a should apply to all epidemics where the autocorrelation function is given by a negative exponential function of distance

(equation 12). Other quadrat-based statistics such as the quadrat variance, σ_r^2 , and the index of dispersion, ID, can be written as explicit functions of quadrat size and κ (equation 1). Thus, equation 12 also allows us to express these variables explicitly in terms of $\rho(0)$, λ , and L . All of the plots based on square quadrats presented by Ridout and Xu (19) of the quadrat variance, the inter-quadrat-correlation, or the index of dispersion versus quadrat size are easily generated using equation 12 and the appropriate values of $\rho(0)$, λ , and L . The derivation of equation 12 is not limited to binary data. The continuous nature of the assumed radial distribution of susceptible plant pairs and the autocorrelation are equally applicable to continuous disease severity data. Indeed, equation 12 is the result of pure geometry as one fits a round peg in a square hole (a radially symmetric correlation function averaged over a square quadrat) and independent of the nature of the data used to calculate variance and spatial correlation.

The use of quadrat-count variation as a measure of aggregation is nothing more than a weighted average of the plant-plant correlation coefficient over the size and shape of a quadrat (equation 12). As quadrat size increases, this weighting considerably reduces the contribution of spatial correlation at distances smaller or larger than half the quadrat size and totally ignores spatial correlation between quadrats. The last difficulty, the inter-quadrat correlation, can be addressed by the use of hierarchical methods (13,14,17,20–23). The approach is to examine the spatial correlation of quadrat counts at length scales larger than a quadrat. This will, of course, provide spatial information at these larger length scales. However, once the data obtained from within a quadrat are pooled, no amount of analysis can recapture the small-scale ($r \ll L/2$) correlations.

Recent orchard-based studies of the spatial analysis of diseases have used a series of different-sized quadrats ($L = 2, 4$, and 8 [literature citation 1] and $L = 2, 3$, and 4 [literature citation 4]). These studies present ample experimental evidence of the strong dependence of the estimated value of κ on the value of L . On the other hand, in both of these studies there are certain data sets which exhibit relatively invariant values for κ , independent of L . Presumably, in these latter instances, the pattern of disease is fortuitously approximated by Pielou's pattern 2, with the length scale of spatial correlation being much larger than the quadrat size ($\lambda \gg L$).

Fig. 7. Plot of normalized correlation $[\rho(r)/s]$ versus normalized distance (r/λ) for simulated spatial epidemics using a Neyman-Scott cluster process. For clarity only 8 of the 80 sets of epidemic parameters are included in this plot. A simple exponential function ($\rho(r) = 0.5e^{-r/\lambda}$ dashed line) is also plotted for comparison. As s increases, there is a reduced correlation at small length scales ($r < \lambda$) and a slight increase in correlation for $r > \lambda$ due to the effects of saturation. In the above simulations, λ is the focal radius, s is the saturation of individual foci, and r is the distance between plant-pairs.

In conclusion, I have extended Pielou's arguments concerning the role of quadrat size on measures of spatial aggregation to binary data sets. This line of thought naturally leads to a self-similar description of the phenomenon. Furthermore, if autocorrelation ($\rho(r)$) can be expressed as an exponential function of distance, $\rho(r) = \rho(0)e^{-r/\lambda}$, the value of the intra-quadrat correlation, κ , can be expressed as an explicit function of $\rho(0)$, λ , and L , in the continuous limit.

As the above analyses show, aggregation, by nature, is not a scalar quantity and cannot be totally characterized by a single measure. Thus, quadrat-based statistics using a single-sized quadrat cannot unambiguously determine spatial aggregation. For this reason, any conclusion based on data taken using one particular quadrat size is useless for comparison, since it is impossible to simultaneously determine both intra-quadrat correlation (κ) and the relative length scale ($a = L/\lambda$) using one value for L . It seems plausible that data from multiple-sized quadrats can be used in a regression analysis based on the relation expressed in equation 12 to obtain simultaneous estimates of κ and L/λ . The determination of these two measures using such an analysis will be left for future work.

APPENDIX

Simulations. The plan is to randomly generate a fixed number, F , of disease foci within the field. Each of these foci is then used to generate infective offspring that are dispersed according to a given radial transport function, $T(r)$, of distance r from center of each focus. For the simulations that follow, I chose the Bessel distribution suggested by van den Bosch et al. (24):

$$T(r) = \frac{1}{2\pi r \lambda} e^{-r/\lambda} \quad (\text{A1})$$

where λ is the characteristic length scale of dispersal. Simulations

are run until I plants are infected within a region containing N plants (8), so that the total probability of infection, p , is given by

$$p = \frac{I}{N} \quad (\text{A2})$$

and, on average, there are i_F infected plants per focus, where

$$i_F = \frac{I}{F} \quad (\text{A3})$$

The mean area of a focus is given by $\pi\lambda^2$, which roughly corresponds to the mean number of plants (healthy and infected) within a focus. Thus, a natural measure of the intensity of a binary epidemic is given by the average disease saturation, s , within a distance, λ , of each disease focus:

$$s = \frac{i_F}{\pi\lambda^2} = \frac{I}{\pi\lambda^2 F} \quad (\text{A4})$$

Since we are interested in patterns exhibiting aggregation, the disease saturation within a focus, s , must be larger than the overall probability of infection, p (i.e., $s > p$). With this in mind, I define the intensity of focal aggregation, $\aleph$, by

$$\aleph = \frac{s-p}{1-p} \quad (\text{A5})$$

which varies from 0 to 1 as foci become increasingly saturated.

For the following simulations, I used a square area with 200 plants on a side. Samples, however, were only taken from an inner 160×160 square area to eliminate edge effects.

The binomial and beta-binomial distributions. Assume that the probability of infection is p within a quadrat containing n plants. Then the probability that i plants are infected is equal to the number of ways to choose i plants out of n plants multiplied by the probability that any given i plants are infected (p^i) multi-

Fig. 8. Plot of the intra-quadrat correlation, κ , calculated using equation 1 (based on the variance) versus κ calculated using equation 9 (based on the correlation function) (Fig. 7) for simulated spatial epidemics using a Neyman-Scott cluster process on a square grid of plants. L is the length of the side of a square quadrat in units of adjacent interplant distance. For each value of five values for L , there are 80 different combinations of epidemic parameters ($p = 0.075, 0.15, 0.3$, and 0.6 ; $\aleph = 0.2, 0.4, 0.6, 0.8$, and 1 ; and $\lambda = 2, 4, 8$, and 16), where p = fraction of plants infected, s = saturation of individual foci, λ is the focal radius, and $\aleph = (s - p)/(1 - p)$ = the intensity of focal aggregation.

plied by the probability that the remaining plants are healthy, $(1-p)^{n-i}$:

$$P_B(i, n, p) = \frac{n!}{i!(n-i)!} p^i \cdot (1-p)^{n-i} \quad (\text{A6})$$

where ! represents the factorial of a number, i.e., $x! = x \cdot (x-1) \cdot \dots \cdot 2 \cdot 1$. Equation A6 defines the binomial distribution (page 148 in literature citation 5) with expectation value for i equal to np and variance about this value equal to $np(1-p)$.

For the case when the probability of infection is not a constant for each quadrat, the binomial distribution is no longer applicable (11). One way of dealing with this variable probability is to assume that the probability of infection for j th quadrat is given by a random variate, P_j , which varies between 0 and 1 and has an average value of p . The choice of the probability density distribution governing the selection of the P_j 's is somewhat arbitrary. However, the beta probability density function, $\beta(P_j)$, is very flexible and has been shown to be consistent with a power law relationship between mean and variance (10):

$$\beta(P_j, p, \kappa) = \frac{P_j^{A-1}(1-P_j)^{B-1}}{Be(A, B)} \quad (\text{A7})$$

where

$$A = \left[\frac{(1-\kappa)p}{\kappa} \right], \text{ and } B = \left[\frac{(1-\kappa)(1-p)}{\kappa} \right]$$

and $Be(A, B)$ is the beta function defined as

$$Be(A, B) = \int_0^1 (1-y)^{A-1} y^{B-1} dy = \frac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)} \quad (\text{A8})$$

where $\Gamma(\cdot)$ is the gamma function (page 47 in literature citation 6). Compounding equation A7 with equation A8 yields the beta-binomial distribution

$$P_{BB}(i, n, p, \kappa) = \frac{n!}{i!(n-i)!} \frac{Be(A+i, B+n-i)}{Be(A, B)} \quad (\text{A9})$$

where A and B are defined as in equation A7. The expectation value for i is np as for the binomial distribution (equation A6) and the variance about this value, σ_i^2 , is given by

$$\sigma_i^2 = np(1-p) \cdot [1 + \kappa(n-1)] \quad (\text{A10})$$

Note that for $\kappa = 0$, the variance approaches the binomial variance.

Pielou's exceptions. Pattern 1. Consider a population of well-separated small foci of disease so widely spaced that they are rarely cut by the edges of a quadrat and do not interfere with one another. In this case the number of infected plants in the j th quadrat is simply the product of the number of foci within the quadrat, f_j , and the number of infected plants within each focus of disease, i_F . Since there are n_j plants in the j th quadrat, assuming foci do not overlap, we can only fit $f_{\max} = n_j/i_F$ foci within the quadrat and the expectation value of f_j is equal to $n_j p/i_F$. In analogy with equation A6, I assume that the probability that f_j foci are in the j th quadrat is given by the following binomial distribution:

$$P_B(f_j, f_{\max}, p) = \frac{f_{\max}!}{f_j!(f_{\max}-f_j)!} p^{f_j} (1-p)^{f_{\max}-f_j} \quad (\text{A11})$$

By definition the variance in f_j is given by $f_{\max} p(1-p)$. Since each focus contains i_F diseased plants, the variance in i is given by

$$\sigma_i^2 = i_F^2 f_{\max} p(1-p) = i_F n_j p(1-p) \quad (\text{A12})$$

Dividing both sides of equation A12 by the binomial variance we see that, for this case, the index of dispersion, $ID = \sigma_i^2/[np(1-p)]$, is independent of quadrat size, n , and numerically equal to the number of diseased plants per focus, i_F .

Pattern 2. Next, consider a mosaic pattern comprised of relatively large patches (much larger than a quadrat) over which disease incidence is relatively constant. In this case, we assume that the patches are so much larger than the quadrats that we can neglect quadrats that straddle the boundaries. Assume that there are M patches and let the number of plants in j th patch be m_j and the probability of infection in the j th patch be p_j . In addition, let the overall probability of infection be p . The total field of study containing N plants is conceptually divided into N/n quadrats with n plants per quadrat, so that the probability that a quadrat lies within the j th patch is given by m_j/N . Then the probability that a quadrat contains i infected plants is given by

$$P(i) = \sum_{j=1}^M \frac{m_j}{N} P_B(i, n, p_j) \quad (\text{A13})$$

where $P_B(i)$ is the binomial distribution (equation A6). By definition the expectation value for i , $E(i)$, is given by

$$E(i) = \sum_{j=1}^M \frac{nm_j p_j}{N} = np; \text{ where } p = \sum_{j=1}^M \frac{m_j p_j}{N} \quad (\text{A14})$$

and the variance σ_i^2 is calculated as

$$\sigma_i^2 = \sum_{i=0}^n (i-np)^2 P(i) = \sum_{i=0}^n \sum_{j=1}^M [(i-np_j) + (np_j-np)]^2 \frac{m_j}{N} P_B(i, n, p_j) \quad (\text{A15})$$

Expanding the quadratic term in equation A15 and interchanging the order of the summations after a bit of manipulation yield

$$\sigma_i^2 = np(1-p) \cdot \left[1 + (n-1) \frac{\sigma_p^2}{p(1-p)} \right] \quad (\text{A16})$$

where

$$\sigma_p^2 = \sum_{j=1}^M \frac{m_j}{N} (p_j - p)^2$$

is the variance in the p_j values.

Intra-quadrat point-pair radial probability distribution [$\phi_L(r)$]. Consider a 2×2 quadrat consisting of four plants ($L = 2$, $n = 4$). Each of the four plants has three neighbors. This makes for a total of $12 = 4 \times 3$ possible pairs of plants. However, in this accounting each pair is counted twice, once for each of its constituent plants. Thus, the number of unique plant-pairs is $n(n-1)/2 = 6$ with $n = 4$. Two of these connections are diagonal across the square and the other four are in line, either vertically or horizontally. The center-to-center distance for plant-pairs in a line is 1 and for diagonal pairs it is $2^{1/2} = (1^2 + 1^2)^{1/2}$. Define the intra-quadrat point-pair radial probability distribution, $\phi_L(r)$, to be the fraction of the total number of plant-pairs from within a square quadrat with L plants on a side, which are a certain distance r apart. For the above 2×2 example: $\phi_2(1) = 2/3$ and $\phi_2(2^{1/2}) = 1/3$. Note that the sum of ϕ over all possible distances is always unity. The same procedure applied to a 3×3 quadrat with $n = 9$ yields 36 plant-pairs ($= 9 \times 8/2$) at five possible distances giving

$$\phi_3(1) = \frac{12}{36}, \quad \phi_3(\sqrt{2}) = \frac{8}{36}, \quad \phi_3(2) = \frac{6}{36}, \quad \phi_3(\sqrt{5}) = \frac{8}{36}, \quad \phi_3(\sqrt{8}) = \frac{2}{36}$$

where the r values are the Pythagorean distances for all possible integral displacements along the x and y axis. The $(\Delta x, \Delta y)$ displacements for each value of r are shown in Table 1. In general, for an $L \times L$ quadrat with $n = L^2$ plants there are $L(L+1)/2 - 1$ unique radial distances given by

$$r_{jk} = \sqrt{j^2 + k^2} \text{ where } j = 1, 2, \dots, L-1 \text{ and } k = 0, 1, \dots, j \quad (\text{A17})$$

and the value for $\phi_L(r_{jk})$ is given by (7)

$$\phi_L(r_{jk}) = \frac{2 \cdot \alpha \cdot (L-j) \cdot (L-k)}{L^2(L^2-1)} \text{ where } \alpha = \begin{cases} 2 & \text{if } j = k \\ 2 & \text{if } k = 0 \\ 4 & \text{otherwise} \end{cases} \quad (\text{A18})$$

where the factor α in equation A18 accounts for the redundancies due to symmetry (Table 1). The above argument can easily be extended to rectangular quadrats ($L \times W$; $L \neq W$) by simple induction. Using equation A18, equation 8 can be rewritten as

$$\kappa = \sum_{j=1}^{L-1} \sum_{k=0}^j \phi_L(r_{jk}) \rho(r_{jk}) \quad (\text{A19})$$

For Pielou's pattern 1 (described previously), the value of ρ in equation A19 would be close to zero except at very small distances ($r_{jk} \ll L$). Thus, k and j would both be much less than the quadrat size ($k, j \ll L$). In this case equation A18 would reduce to some constant divided by L^2 (equation 2). For Pielou's pattern 2, the value of ρ would be a constant and so the value of κ would be relatively independent of quadrat size. To further investigate the behavior of the ϕ_L 's (equation A18), I have calculated the first and second spatial moments for selected values of L in Table 2. When the moments are scaled to the length scale of the quadrat (Table 2, last three columns), we see that the mean and the standard deviation approach asymptotic limits of $L/2$ and $L/4$, respectively. Due to the discrete nature of the ϕ_L 's and the irregular spacing of the r_{ij} 's, plots of these functions are not very informative. For this reason, I define the cumulative pair distribution, $\Phi_L(r)$, to be

$$\Phi_L(r) = \sum_{r_{ij} \leq r} \phi_L(r_{ij}) \quad (\text{A20})$$

which is plotted as a function of r/L for $L = 2, 4, 8, 16$, and 32 in Figure 6. In the asymptotic limit ($L \rightarrow \infty$) the shape of this curve is equivalent to the edge-effect correction factor for "Ripley's L " in a square field derived elsewhere for $r < L/2$ (25).

In the asymptotic limit ($L \rightarrow \infty$), the discrete function, $\phi_L(r_{jk})$, given by equation A18 becomes the continuous radial density function, $\phi(r/L)$, defined to be the fraction of the total number of intra-quadrat distances that lie between r/L and $r/L + dr/L$, given by

$$\phi_L(r) = \frac{1}{L^3} \int_0^{2\pi} \alpha(\theta, r) \cdot (L - |r \sin \theta|) \cdot (L - |r \cos \theta|) \cdot r d\theta \quad (\text{A21})$$

where

$$\alpha(\theta, r) = \begin{cases} 1; & r \leq L \\ 1; & r > L, \quad r \sin \theta \leq L, \quad r \cos \theta \leq L \\ 0; & r > L, \quad r \sin \theta > L, \quad r \cos \theta > L \end{cases}$$

The rather complicated behavior of the function $\alpha(\theta, r)$ in equation A21 accounts for the corners of the square quadrat. Equation A21 can be expanded to yield

$$\phi_L(r) = \int_0^{2\pi} \alpha(\theta, r) \cdot \left(1 - \frac{r}{L} |\sin \theta| - \frac{r}{L} |\cos \theta| + \left(\frac{r}{L} \right)^2 |\sin \theta| \cdot |\cos \theta| \right) \cdot \frac{r}{L} d\theta \quad (\text{A22})$$

Taking advantage of symmetry and setting $R = r/L$, we obtain

$$\phi_L(r) = \begin{cases} 8 \cdot \int_0^{\pi/4} (1 - R \sin \theta - R \cos \theta + R^2 \sin \theta \cdot \cos \theta) \cdot R d\theta & ; r \leq L \\ 8 \cdot \int_{\cos^{-1}(1/R)}^{\pi/4} (1 - R \sin \theta - R \cos \theta + R^2 \sin \theta \cdot \cos \theta) \cdot a d\theta & ; r > L \end{cases} \quad (\text{A23})$$

where the lower limit in the second integral in equation A23 reflects the functional form of $\alpha(\theta, r)$ presented in equation A21. The above integrals can be easily evaluated to yield

$$\phi(r/L) d(r/L) = \begin{cases} \frac{r}{L} \left[2\pi - 8\left(\frac{r}{L}\right) + 2\left(\frac{r}{L}\right)^2 \right] & ; r \leq L \\ \frac{r}{L} \left[2\pi - 4 - 8\cos^{-1}\left(\frac{r}{L}\right) + 8 \cdot \sqrt{\left(\frac{r}{L}\right)^2 - 1} - 2\left(\frac{r}{L}\right)^2 \right] & ; r > L \end{cases} \quad (10)$$

The above result is identical to that obtained assuming a homogeneous distribution of foliage throughout the quadrat.

If we now assume that the correlation function can be expressed as a negative exponential function of length scale, λ , equation A19 can be rewritten as

$$\kappa = \frac{1}{L} \int_0^{\sqrt{2}L} \phi\left(\frac{r}{L}\right) \rho(r) dr \quad (11)$$

where

$$\rho(r) = \rho(0) e^{-r/\lambda}$$

The integral in equation 11 can be broken down to two integrals: one ranging from $r = 0$ to L and the second from L to $2^{1/2}L$. The first integral using the form for $\phi(r/L)$ appropriate for $r \leq L$ (equation 10) is easily evaluated in closed form (first four terms in equation 12) and accounts for more than 97.5% of the value of κ . The second integral presents a bit more difficulty and the resulting closed form solution is rather complex. However, a simple Taylor series expansion about $r = 2^{1/2}L$ yielded a numerical approximation to the second integral accurate to third order in the quantity, $2^{1/2} - R$ (error $< 0.002 \kappa$, fifth term in equation 12).

LITERATURE CITED

- Bassanezi, R. B., Bergamin Filho, A. B., Amorim, L., Gimenes-Fernandes, N., Gottwald, T. R., and Bové, J. M. 2003. Spatial and temporal analyses of citrus sudden death as a tool to generate hypotheses concerning its etiology. *Phytopathology* 93:502-512.
- Cochran, W. G. 1977. *Sampling Techniques*. John Wiley, New York.
- Cressie, N. 1991. *Statistics for Spatial Data*. (Pages 313-337 in section 5.6, Spatial design: Pages 661-669 in section 8.5.3, Poisson cluster process). John Wiley & Sons, New York.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J.-B. 2003. Spatial pattern analysis of Sharka disease (*Plum pox virus* strain M) in peach orchards of southern France. *Phytopathology* 93:1543-1552.
- Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, Vol. 1. 3rd ed. John Wiley, New York.
- Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, Vol. 2. 3rd ed. John Wiley, New York.
- Ferrandino, F. J. 1996. Two-dimensional distance class analysis of disease-incidence data: Problems and possible solutions. *Phytopathology* 86:685-691.
- Ferrandino, F. J. 2004. Measuring spatial aggregation in binary epidemics: Correlative analysis and the advantage of fractal-based sampling. *Phytopathology* 94:1215-1227.
- Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. *Plant Pathol.* 41:657-660.
- Hughes, G., and Madden, L. V. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. *Phytopathology* 83:759-763.
- Hughes, G., and Madden, L. V. 1994. Intracluster correlation and the problem of sample size determination. (Abstr.) *Phytopathology* 84:1080.
- Hughes, G., Madden, L. V., and Munkvold, G. P. 1996. Cluster sampling for disease incidence data. *Phytopathology* 86:132-137.
- Hughes, G., McRoberts, N., Madden, L. V., and Gottwald, T. R. 1997. Relationships between disease incidence at two levels in a spatial hierarchy. *Phytopathology* 87:542-550.
- Hughes, G., Gottwald, T. R., and Levy, L. 2002. The use of hierarchical sampling in the surveillance program for *Plum pox virus* incidence in the United States. *Plant Dis.* 6:259-263.
- Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distributions, heterogeneity, and temporal analysis. *Annu. Rev. Phytopathol.* 33:529-564.
- Neyman, J., and Scott, E. L. 1958. Statistical approach to problems in cosmology. *J. Roy. Stat. Soc. B.* 20:1-29.
- Ngugi, H. K., King, S. B., Abayo, G. O., and Reddy, Y. V. R. 2002. Prevalence, incidence, and severity of sorghum diseases in western Kenya. *Plant Dis.* 86:65-70.
- Pielou, E. C. 1969. *An Introduction to Mathematical Ecology*. John Wiley & Sons, New York.
- Ridout, M. S., and Xu, X.-M. 2000. Relationships between several quadrat based statistical measures used to characterize spatial aspects of disease incidence data. *Phytopathology* 90:568-575.
- Scott, J. B., Hay, F. S., Wilson, C. R., Cotterill, P. J., and Fitt, A. J. 2003. Spatiotemporal analysis of epiphytotic of downy mildew of oilseed poppy in Tasmania, Australia. *Phytopathology* 93:752-757.
- Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis and sequential sampling for the incidence of leaf spot on strawberry in Ohio. *Plant Dis.* 83:992-1000.

22. Turechek, W. W., and Madden, L. V. 2000. Analysis of the association between the incidence of two spatially aggregated foliar diseases of strawberry. *Phytopathology* 90:157-170.
23. Turechek, W. W., and Madden, L. V. 2003. A generalized linear modeling approach for characterizing disease incidence in a spatial hierarchy. *Phytopathology* 93:458-466.
24. van den Bosch, F., Zadoks, J. C., and Metz, J. A. J. 1988. Focus expansion in plant disease. I: The constant rate of focus expansion. *Phytopathology* 78:54-58.
25. Ward, J. S., and Ferrandino, F. J. 1998. New derivation reduces bias and increases power of Ripley's L index. *Ecol. Model.* 116:225-236.
26. Xu, X.-M., and Ridout, M. S. 1998. Effects of initial epidemic conditions, sporulation rate, and spore dispersal gradient on the spatio-temporal dynamics of plant disease epidemics. *Phytopathology* 88:1000-1012.
27. Xu, X.-M., and Ridout, M. S. 2000. Effects of quadrat size and shape, initial epidemic conditions, and spore dispersal gradient on spatial statistics of plant disease epidemics. *Phytopathology* 90:738-750.